1. Molecular Machines: Protein Complexes

SECTION

The human genome is a sophisticated and complex coding system for the formation of thousands of different proteins. Complex assemblies of these proteins, the molecular machines of the cell, perform the majority of cellular functions, from enzymatic activity to cell motility. Goal 1 of the Genomes to Life (GTL) program is to "identify and characterize the molecular machines present in the microbial cell." This goal is intricately linked to the ultimate GTL program goal: to model cells and cell systems using a predictive, computational biology approach for understanding cellular behavior. A key to developing these modeling capabilities is the ability to produce comprehensive functional characterization of gene products acting in concert in large, multiprotein molecular machines. Understanding the constituents and functions of the machines is critical to the overall goals of the program (Figs.1.1a and 1.1b).

A comprehensive approach to molecular machines requires the identification of all molecular machines in the cell, their protein constituents, the interactions that bind the constituents together, and how these machines are assembled, regulated by their environment, and perform their function. This effort will be aided by comparative genomics among species to furnish information critical to the identification of novel molecular machines. The size of the task should not be underestimated. The active and existing complexes in any given cell depend on (1) the genome, (2) the cellular environment, and (3) time. Key to progress will be identification of the complexes most important to understanding cell function. Imaging will play a key role in this process.

Issues and Limitations

Proteins and other molecules that comprise molecular machines may be identified through the mapping of protein-protein and protein-ligand

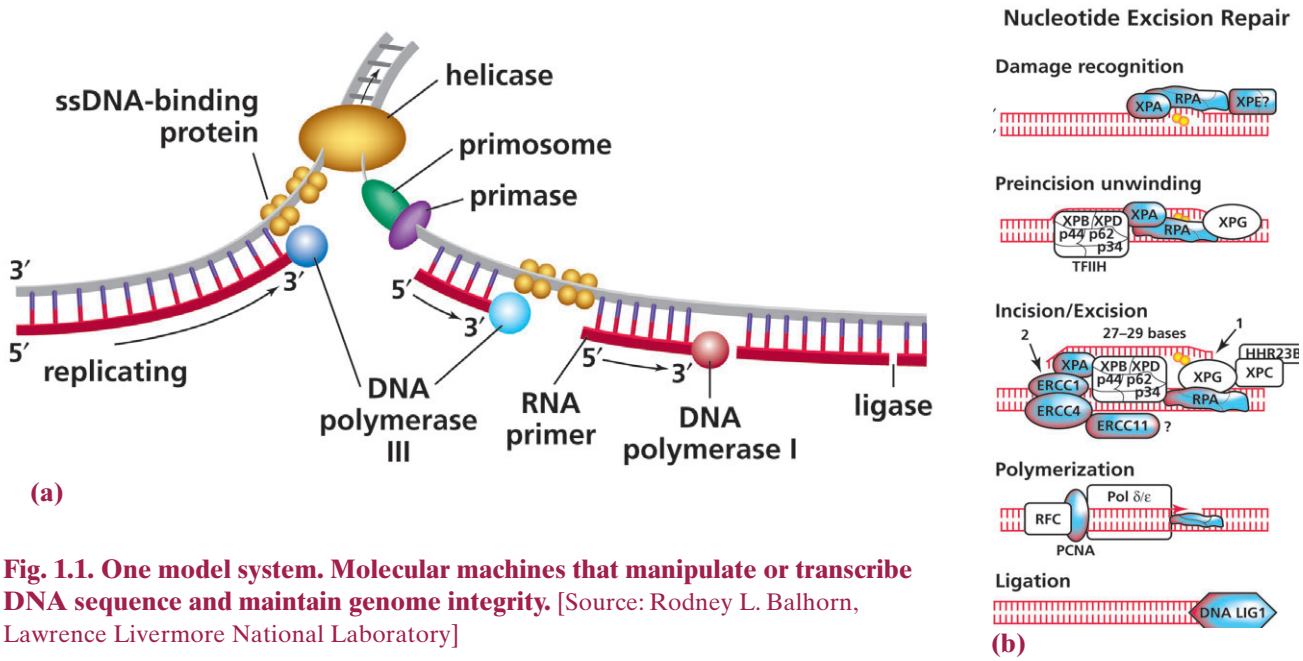


Fig. 1.1. One model system. Molecular machines that manipulate or transcribe DNA sequence and maintain genome integrity. [Source: Rodney L. Balhorn, Lawrence Livermore National Laboratory]

interactions using phage display, two-hybrid techniques, and direct isolation of protein complexes. These methods allow independent enumerations of intermolecular interactions, and correlations derived by comparing results from different functions will provide significant information for elucidating signaling pathways and the many other functions performed by molecular machines; each method, however, has intrinsic biological limitations. These methods need optimizing (by other parts of the GTL program) so they can deliver a more rapid identification of the constituents of molecular machines. This is a significant need that will directly support future imaging efforts.

Required Instruments and Methods

A molecular-level understanding of a molecular machine's function requires imaging of the machine at various levels of detail using techniques that measure many variables (e.g., chemistry and structure) and enable collection of dynamic information on numerous time scales. Imaging of individual proteins and their complexes will be required to

- Provide a molecular-level understanding of their function through direct observations,
- Test hypotheses about their function and regulation through designed mutations, and
- Develop probes that will enable the visualization and tracking of individual molecular machines inside living cells.

In general, this will require the coordinated use of multiple imaging techniques. The goal will be to monitor structure, function, stimuli, and response in real time; for example, the onset of a change in calcium ion concentration while simultaneously observing calcium-induced changes in molecular structure.

Ultimately, the molecular machine must be imaged in the context of its cellular environment to understand where and when it functions and to identify other cellular components or systems that may interact with it during its functioning. As a starting point, the intact machine must be imaged when isolated from other cellular components to define the machine's extent. Machine component positions may be determined by using affinity tags (such as antibodies or antibody domains) specific to individual components. Machine components (e.g., constituent proteins) may need to be imaged individually at very high resolution. Images of dynamic activities, which can provide considerable insight into the machine's function and molecular mechanisms, should be obtained wherever possible.

Imaging at the whole-cell level is discussed in detail in other parts of this report but may be carried out, for example, using optical microscopy (e.g., fluorescence microscopy, Fig. 1.2), 2D or 3D confocal microscopy, X-ray microscopy (2D or 3D X-ray tomography), or electron microscopy (2D or 3D electron tomography). Optical microscopy provides images of living cells and enables the collection of time-domain information. Three-dimensional scanning multiphoton microscopy can provide 3D images of the cell's dynamic behavior. However, significant advances in its time resolution would greatly advance our ability to study the activity of living cells in great detail.

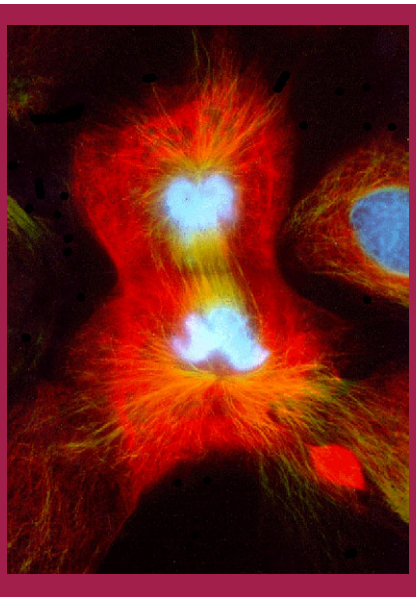


Fig. 1.2. Imaging at multiple-length scales using different techniques to determine the function of molecular machines. This triple-exposure fluorescent micrograph is of a newt lung cell in early telophase of mitosis. The red stain is keratin, and the yellow is tubulin; blue depicts chromosomes. [Image courtesy of Conly L. Rieder; Division of Molecular Medicine; Wadsworth Center; New York State Department of Health; Albany, New York]

Protein Identification and Characterization

Protein identification and characterization are currently slow, labor-intensive processes. New, more rapid methods are needed that will enable identification of the protein constituents comprising molecular machines, types and location of post-translational modifications, and ligands that participate in their function. More efficient protein-sequencing methods will enable identification of the protein, and, when coupled with mass spectrometry, the sites of post-translational modifications.

Current Methods

Affinity tags (e.g., antibodies) against a molecular machine will be needed to (1) enable isolation of the entire molecular machine, as well as its component proteins, and (2), when combined with imaging, enable localization of both the molecular machine within the cell and a particular constituent within the machine. Affinity tags can be produced using a variety of techniques, among them phage and yeast display; however, advances in high-throughput approaches are needed for wholesale production of affinity tags for the cell's many molecular machines.

New Approaches

In addition to developing new probes for tagging proteins in vivo, new methods are needed that will enable the tagging of proteins at a single site within their sequence in vivo with a chromophore, electron-dense material, or other suitable probes. This may include the development of new vectors or gene constructs that will result in the expression of proteins covalently attached to unique, modifiable functional groups (e.g., thioester) or relatively short peptide sequences that preferentially (and when possible, covalently) bind fluorescent ligands (e.g., metals, lanthanides, and particles) or contribute their own signal.

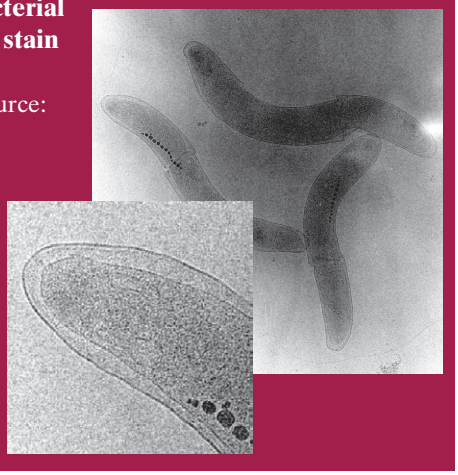
Imaging Technology Overview

A complete understanding of molecular machines will require information covering a wide range of dimensions and time scales. Structural data such as molecular dimensions, subunit structure, and overall molecule shape are typically obtained with high-resolution techniques at the angstrom to nanometer scale. Information on dynamics and function of molecular machines, however, is usually obtained on larger spatial and shorter time scales using methods that interrogate a molecular machine within its immediate environment. A variety of techniques can be used to access the structural and dynamic information of protein complexes. However, there is much room for improvement in several areas, including the possibility of exploiting new phenomena.

Short-Wavelength Techniques

Electrons and X rays both have wavelength ranges that can give diffraction-limited resolution in a range useful for directly imaging the spatial arrangement of subunits in isolated protein complexes. X-ray crystallography has provided atomic-resolution structures of proteins, their subunits, and small machines such as ribosomes. Electron microscopy (EM) is well established for viewing complexes at 0.5- to 2.0-nm resolution. This resolution is useful for direct structure determination or for positioning subunit structures from X ray into an overall atomic-resolution complex. Samples are rapidly frozen and viewed in thin amorphous ice. Particles are typically viewed by tomography either in 2D arrays or as isolated objects. If the sample is freeze-dried, quantitative measurement of electron scattering can give direct mass distribution useful for determining subunit stoichiometry or confirming the faithfulness of reassembly (Fig. 1.3).

Fig. 1.3 *Magnetospirillum magnetotacticum*. Electron microscopy of intact, frozen-hydrated bacterial cells with no stain or chemical fixatives. [Source: Kenneth H. Downing, Lawrence Berkeley National Laboratory]



Intact molecular machines composed of multiple proteins often are too large or too flexible to make a crystallographic approach feasible. Furthermore, to obtain any information about substructure, resolutions significantly greater than that of light microscopies are necessary. X-ray microscopy is promising for the future when improvements in optics and instrumentation will take full advantage of the 2.4-nm wavelength used for image formation to yield fine details of protein complexes. At present, this size is best suited to EM.

The imaging of flash-frozen specimens of molecular machines has huge potential for providing large amounts of structural and functional information about the operation of molecular machines. Flash-frozen specimens maintain their native conformation and can survive multiple exposures so that a partial tilt series of micrographs can be produced on a single specimen. Intermediate voltage electron micrographs (300 to 400 keV) are particularly well suited to this type of study because these relatively high energy electrons have a greater penetration depth and cause less radiation damage than the more typical 100-keV electrons used in most transmission electron microscopes (TEM). Three-dimensional reconstructions can be calculated from images of a molecular machine taken at many different angles (much like the production of 3D radiological images from a CAT scan). Images of affinity-tagged specimens give significant information about the relative positioning of different gene products within the machine (Figs. 1.4 and 1.5).

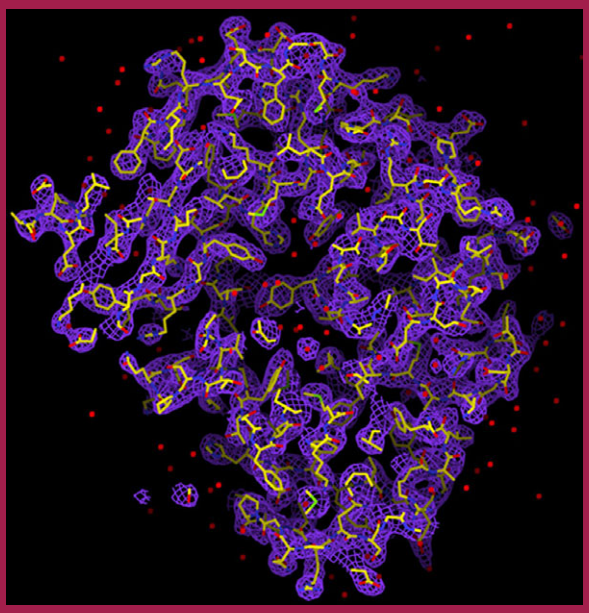


Fig. 1.4. Three-dimensional structural determination of single rubisco molecules with a simulated X-FEL and direct-phase retrieval by oversampling technique. The map shows the 3D electron-density map of the rubisco molecule (contoured at 2 sigma) reconstructed from the 3D diffraction pattern of 3×10^5 identical copies of the rubisco molecule. Poisson noise was added (Rl – 16.6%), and the central 3 × 3 × 3 pixel intensity was removed. The edge of the simulated X-FEL diffraction pattern corresponds to a resolution of 2.5 Å. [Source: Jianwei Miao, Keith Hodgson, and David Sayre, "An Approach to Three-Dimensional Structures of Biomolecules by Using Single-Molecule Diffraction Images," *PNAS* **98**(12), 6641–45. ©2001 National Academy of Science, U.S.A.]

Theoretical studies of the number of images needed to produce very high resolution images of a molecular machine have been put forward by a number of investigators. Theoretically, a 3D reconstruction of a molecular machine can be produced at a level of resolution approaching that of X-ray crystallography.

This process is not instrument limited but rather computation limited. Estimates predict that an image with 4-Å resolution of the ribosome's large subunit could be computed from 500,000 particle images and an image with 3-Å resolution from 2 million particle images. Even though hundreds of particle images can be collected on one electron micrograph, with CCD detectors rather than film, the ability to process and recombine this

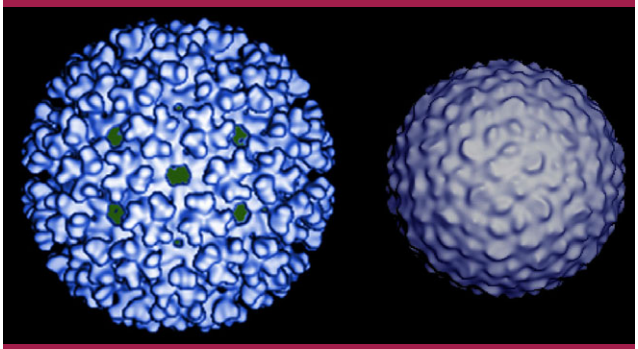


Fig. 1.5. A 3D reconstruction from electronmicrographs of flash-frozen specimens of the Ross River virus (left) and the Dengue 2 virus. [Source: Tim Baker and Richard Kuhn, Purdue University]

number of images is still far beyond the capabilities of existing algorithms and computational resources. The huge potential of this approach, however, could justify a significant investment in research directed toward achieving that capability.

Scanning transmission electron microscopy (STEM) provides information complementary to TEM techniques by showing a direct link to the biochemistry. Measurement of mass-per-unit length of helical complexes and total mass of other complexes supplies important information for determining the stoichiometry of the complexes and comparing native and reassembled structures. Use of metal clusters such as undecagold as labels for specific constituents makes STEM a valuable method for localizing gene products within complexes (Fig. 1.6).

Imaging molecular machines within cells can be achieved using soft X-ray (XM) or EM. High-resolution techniques such as XM can be carried out at atmospheric pressure on whole hydrated cells up to 10 μm thick and can obtain information at 25- to 30-nm resolution using existing optics.

Better resolution can be obtained using TEM of thin specimens placed in a vacuum chamber and held stationary. A promising method for preserving specimens for both XM and TEM is flash-freezing, in which the specimen is frozen so quickly that water molecules do not have time to crystallize but instead solidify into a vitreous ice. This well-developed technology enables cryo-X-ray tomography of whole cells yielding structural information at better than 50-nm resolution and electron tomography of isolated complexes and 2D arrays approaching 1-nm resolution (Fig. 1.7).

The resolution limit of electron tomography is not that of the microscope but is related to the obtainable number of distinct views of the cell. Because every cell is different, images from different cells cannot be averaged (as is typically done with images of molecular complexes). Instead, an entire tilt series must be carried out on a single, small cell (e.g., bacteria <0.5 μm in diameter) or on sections of cells. Micrographs resulting from the bacterium's tilt series or of all sections of a eukaryotic cell can then be recombined into a 3D reconstruction approaching 10 nm in resolution (Fig. 1.8). Methods for lowering radiation damage are needed to allow for larger numbers of images to be obtained on a single specimen. This may be possible using higher-voltage electrons (>500 keV) or by imaging specimens maintained at liquid helium temperatures.

Plastically Stressed Adhesion Pili

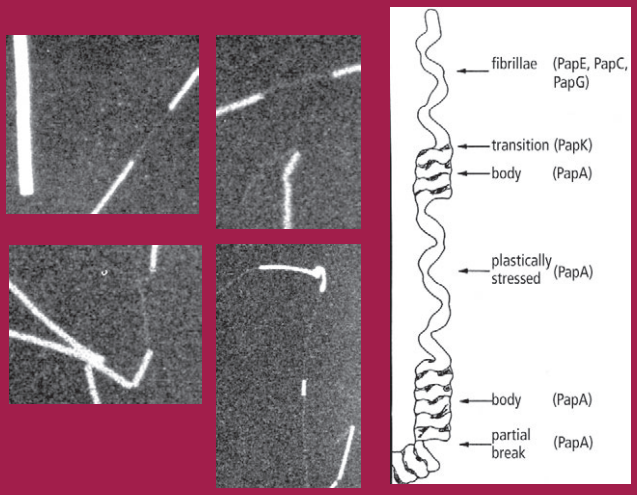


Fig. 1.6. Helical structure of P-pili from *Escherichia coli*, showing evidence from X-ray fiber diffraction and scanning-transmission electron microscopy. [Source: Minfang Gong and Lee Makowski, "Helical Structure of P- pili from *Escherichia coli*: Evidence from X-Ray Fiber Diffraction and Scanning-Transmissions Electron Microscopy," *Journal of Molecular Biology* **228**(3), 735–42 (1992). Reprinted by permission of the publisher Academic Press, an imprint of Elsevier Science]

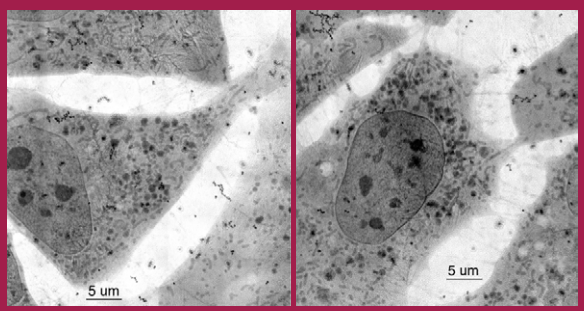


Fig. 1.7. Cryo-X-ray microscopy of 3T3 fibroblast whole cells with no fixatives, stains, or contrast enhancement. [Source: W. Meyer-Ilse et al., "High-Resolution Protein Localization Using Soft X-Ray Microscopy," *Journal of Microscopy* **201**, 395–403 (2001)]

Dynamic information on an individual specimen that cannot be generated with EM can be obtained by flash-freezing a specimen at various times after an applied stimulus. To study structural changes, for instance, Ken Taylor at Florida State University is flash-freezing a specimen of an insect flight muscle in the act of contracting.

Recently, electron microscopes using energy filters have come into use. One potential use of energy filters is the development of energy-specific tags to absorb electrons at one energy but not at others. These tags could be used in a manner analogous to that of green fluorescent protein in light microscopy—to localize tagged gene products within cells at EM resolution. Used with electron tomography, these tags could localize specific molecular machines to within 10 nm in the cellular environment.

Optical Microscopy

Single-molecule microscopy and spectroscopy can give direct insight into the mechanism, dynamics, and kinetics of molecular machines. These molecular complexes are dynamic in their behavior, lending themselves to study at the single-molecule level. This approach can provide a broader view of population distributions than techniques that probe the ensemble average.

Laser-based optical probe techniques show great promise for real-time monitoring of the motions of molecular machines. Techniques are evolving that side-step diffraction limits of focused light and that could monitor individual molecules in vivo and in vitro. Multiphoton and multicolor techniques hold promise for high selectivity and the simultaneous detection of multiple species. Advances in detector technology are still needed to permit the real-time imaging of multiple species, making possible the simultaneous monitoring of chemical composition and molecular motion. Multiplexed measurements performed in real time could uncover details of the mechanisms for building and operating molecular machines.

A highly desirable goal is the capability to directly view and study the workings of molecular machines (i.e., protein complexes). This accomplishment is constrained by the sizes of complexes that range from angstroms to tens of nanometers. The size range is much smaller than the resolution of conventional optical microscopic techniques, which are limited by the diffraction limit of light. Other approaches, such as scanning electron microscopy (SEM), TEM, STEM, and atomic force microscopy (AFM, Fig. 1.9), have the resolution to directly image at these scales; however, these techniques require experimental conditions that do not lend themselves to real-time study of dynamic systems. Strategies for observing the operation, motion, and assembly of a molecular machine as a function of time are possible and should be further developed.

One limitation of current methods is that external tags must be used to observe the structure or motion of the system of interest. These tags must be highly selective, nonperturbative, easily detected, resistant to degradation, and have unique spectroscopic signatures. The most common probes for this purpose are fluorescent markers, even though they do not have optimal properties. Several new types of probes being developed for monitoring molecular motion include phosphorescent structures, quantum dots, and fluorescent microspheres, as well as surface-enhanced Raman particles and plasmon-resonant particles.

Fig. 1.8. Section from a 3D electron tomographic reconstruction of *Caulobacter*. This cell is budding off from the parent cell (right), while developing a flagellum at the left end. The sample was prepared in vitreous ice with no chemical fixatives or stains. Internal features such as membranes are clearly seen, as are protein complexes about the size of ribosomes and patches of the periodic S-layer. [Source: Kenneth H. Downing and Grant J. Jensen, Lawrence Berkeley National Laboratory]

Each probe has trade-offs such as size limitations and optical activity lifetime. Examples of trade-offs include a comparison of single-fluorescent molecule and fluorescent-microsphere tagging approaches. The single-fluorescent molecule tag provides a probe that is small relative to the machine under investigation, which ensures that the probe will not significantly alter the dynamics and kinetics—a concern when the microsphere used is much larger than the system under investigation. However, the single molecule has a much shorter optically active lifetime of milliseconds to minutes compared with the microsphere, which can be active for hours. Additionally, the instrument response of single-molecule emission is complicated by dynamic photophysical behavior. Despite these limitations, tags are currently used to directly observe biomolecular movement at a much higher resolution than the diffraction limit associated with laser focus. This is accomplished by fitting the optical image with the known or estimated instrument-response function. If the fit parameters accurately reflect the response function, the tag location can be determined down to the nanometer level.

Another approach to investigate molecular movement consists of observing an indirect signal that can be correlated with the movement. One such approach is single-molecule fluorescence resonance energy transfer (SM-FRET). This method measures angstrom-scale distances between two fluorescent tags. A fluorescent molecule that has absorbed a photon and becomes electronically excited has several pathways for removing this energy. It can fluoresce, internally relax (internal conversion or intersystem crossing), be quenched, or transfer its excitation to another chromophore. Any of these processes can be sensitive to the molecule's environment, and measurements of their rates can be used to probe that environment. Energy transfer from donor to acceptor chromophore is particularly interesting because of its sensitivity to the distance between chromophores. This energy-transfer behavior is a dipole-dipole interaction and can measure molecular-scale distances (Fig. 1.10).

Currently, the typical FRET experiment determines the degree of energy transfer by measuring the total amount of light detected through two different spectral filters. The average distance between FRET pairs can be calculated from the amount of light originating from each member of the pair. Using FRET on single pairs makes possible the determination not only of the average distance between FRET pairs but also of distance distribution within a sample. Individual molecule-correlation functions can be calculated and averaged to obtain overall correlation functions and therefore the power spectrum. From this information and the quantum yields of donor and acceptor, we can determine energy-transfer efficiency and distances between chromophores. If one of the FRET pair chromophores is attached to one part of a molecular motor and the other to a different part, their relative motions can be monitored.

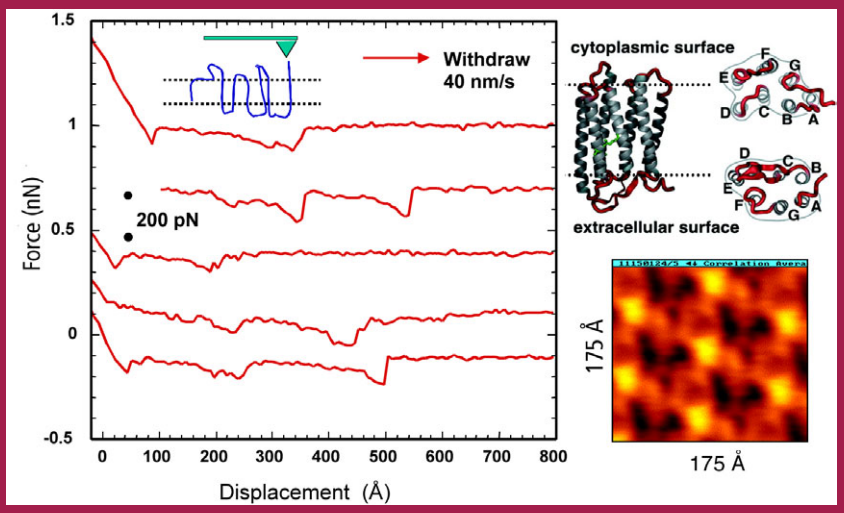


Fig. 1.9. Atomic force microscopy providing structural information on membrane proteins via imaging and single-molecule force curves of protein unfolding. [Source: Alan R. Burns, Sandia National Laboratories]

Other approaches can include the observation of such optical properties as polarization state, fluctuations of excited-state lifetime, and fluorescence depolarization. All require labeling of the molecular complex under investigation, which in some cases is needed in multiple

sites. Another optical approach involves monitoring movement or structural changes based on the intrinsic spectroscopic properties of the system under study. This makes the use of a tag unnecessary. Surface-enhanced Raman scattering is a possible strategy that uses a common spectroscopic property to produce a fingerprint spectrum of the structure and show correlated spectral fluctuations with molecular structure. The use of surface-enhanced approaches and predetermined spectral features correlated with the movement of interest should be capable of millisecond time scales over extended periods.

Scanning Probe Microscopy (SPM)

SPM has proven useful for determining biological structure and function. It can be performed in situ (on live organisms, if necessary) without staining. This technique potentially provides 1-nm topographic resolution. SPM can be combined with and enhance simultaneous fluorescence imaging. It also can be used to image individual proteins, protein complexes, and molecular arrays (membrane proteins). Phase-sensitive tapping mode provides information about chemical changes and "softness" (compliance). Functionalized tips can be used to generate force vs displacement curves, providing the ability to obtain single-molecule bond strength information that lends itself to the characterization of protein-protein or protein-ligand interactions and tension-induced protein unfolding.

Diffraction

Information about the size and shape of purified structures randomly oriented in solution can be obtained by X-ray or neutron scattering. Neutron diffraction is particularly useful for defining the volume of the structure that excludes water through the use of scattering in varying ratios of H_2O and D_2O. Variation of H_2O and D_2O ratios can be used to contrast deuterated proteins, making possible the study of ligand binding, including conformational changes induced by the interaction. Wide-angle solution scattering also has the potential to provide scattering fingerprints that can be used to determine the structural class of a protein in solution. For instance, this technique can readily distinguish between an immunoglobulin fold and a beta-barrel.

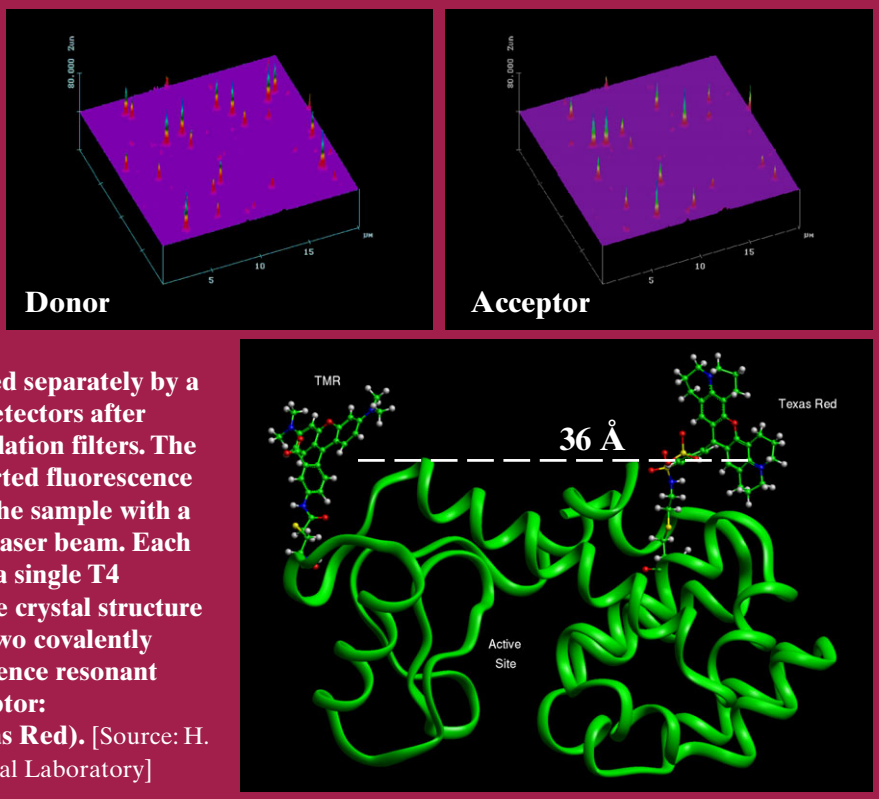


Fig. 1.10. Top: Two-channel fluorescence images (20 μm by 20 μm) of individual donor- or acceptor-labeled T4 lysozyme molecules tethered by cross-linker molecules to the hydrocarbon-modified glass surface of a cover slip under a pH 7.2 aqueous buffer solution. The donor and acceptor emissions were detected separately by a pair of avalanche photodiode detectors after passing through appropriate isolation filters. The images were taken with an inverted fluorescence microscope by raster-scanning the sample with a focused 10- to 100-nW, 532-nm laser beam. Each individual peak is attributed to a single T4 lysozyme molecule. Bottom: The crystal structure of wild-type T4 lysozyme with two covalently labeled dye molecules (fluorescence resonant energy transfer donor and acceptor: tetramethylrhodamine and Texas Red). [Source: H. Peter Lu, Pacific Northwest National Laboratory]

Cross-Cutting Needs

In future we anticipate a need for laboratory-scale sources for X-ray diffraction and electron scattering that will give scientists dedicated instruments for determining the structure of individual proteins and protein complexes comprising a cell's molecular machines. New techniques capable of sequencing single proteins online also will be needed for instant identification of a molecular machine's constituents and its location in the genome. Some of these advances will be enabled by microfluidics and microchemistry lab efforts. We anticipate needing new techniques for the specific labeling of proteins and protein domains as well as for manipulating the genome to better characterize and customize molecular machines.

Other DOE programs could meet several imaging needs. These include new probes, advances in protein sequencing and characterization, sample preparation and validation, computational simulations, and data management.

Rapid Methods for Structure Determination of Single-Protein Molecules or Complexes

Current methods for determining protein structure take years to provide a finished structure. Nuclear magnetic resonance (NMR) structure determination is currently limited to proteins of 40 to 60 kD, and X-ray diffraction methods require that the proteins be crystallized. Crystallization is the slowest step in the process, often taking many years of concerted effort. In addition, many proteins never crystallize, among them membrane proteins. Frequently, molecular signaling transverses the membrane, and more than 30% of all proteins are known to be tightly associated with membranes. New biochemical and biophysical approaches will be required to identify the molecular structures of these proteins. Creating tabletop or lab-scale X-ray sources would be a high-impact development that would allow the scientific community to take full advantage of the use of X rays in imaging protein machines.

New developments leading to the production of a "next-generation" beamline, such as the Linac Coherent Light Source at Stanford, offer the promise of obtaining high-resolution (2 Å or less) diffraction data from single-protein molecules, protein complexes of any size, viruses, and even cells without crystallization using a coherent X-ray source ten times brighter than we have today. To aid the development of this technology, which may become available as early as 2007, single-molecule diffraction techniques using this beamline will need to be established. Other needs include sample preparation and introduction, rapid-readout detection arrays, and computer algorithms for combining and interpreting diffraction data from tens of thousands of molecules.

The development of new molecular probes applies to almost all aspects of GTL imaging-based research. Section 6 summarizes research needs in this area. Identification and characterization of molecular machines will require a wide range of probes. For example, optical methods require probes with high efficiency and stability for single-molecule research as well as probes designed for use with multiphoton imaging methods. Short-wavelength methods require electron-dense probes for use as high-resolution labels for individual protein subunits in complexes.

Protein or Sample Characterization

Development of High-Throughput Methods for Identifying Protein Function

Identifying protein function can be a daunting task. The GTL program would be advanced significantly if methods could be developed to enable the rapid identification of an unknown protein's function. Continued increases in numbers of known (experimentally derived) protein structures with different folds will aid the analysis of protein sequences as a first step to predict function using such computational methods as comparative modeling and threading. Function may be inferred or predicted by identifying unique folds, particular combinations of structural motifs, or structural features that match known substrate, cofactor, or other binding sites.

New array or chip-based protein-function assays should be developed for experimentally screening large numbers of proteins for function or functional attributes. This could be accomplished, for example, using arrays that test proteins for their ability to bind to a wide assortment of small molecules (e.g., metals, cofactors, and carbohydrates), macromolecules (e.g., ssDNA, dsDNA, RNA, and specific DNA sequences), antibodies, or other proteins. Direct assays that actually monitor enzymatic function also could be developed in similar array format.

Specimen Integrity and Functionality

Challenges for the imaging of molecular machines remain significant. First, new methods must be developed for producing enough of these machines for biochemical, biophysical, and imaging studies. In many cases, protein constituents of molecular machines are insoluble on their own in the absence of other machine constituents. Only by coexpression of multiple constituents—ideally all constituents—is it possible to produce specimens containing soluble samples of these machines. Although some coexpression vectors are available, additional ones need to be developed to enable the production of the many types of molecular machines to be studied in the GTL program.

The integrity of the purified molecular machine also must be assured. Integrity depends on the isolation protocol, extra components that may be required for stability, other preparative procedures (e.g., freezing), and deposition on the substrate used for imaging. Ideally, a functional assay should be available to ensure the viability of the complexes as prepared for imaging. Alternately, the consistency of results obtained by different imaging modalities and specimen preparations must be assessed. Fluorescence-based single-molecule (complex) spectroscopy combined with AFM structure mapping has the potential to assay for sample purity and reveal the relationship between secondary structure and dynamics or function.

Computational Simulations

Simulations of the structure, activity, and assembly of a molecular complex will need to be carried out on multiple-length scales and various levels of detail. For instance, the most detailed simulations will involve molecular models including atomic coordinates of all constituents. Subunit structures from X-ray crystallography may be packed together to provide atomic-resolution models. Less-detailed models will be useful for simulating the dynamics of complex formation, mechanical-force generation, and docking. Other models will be needed to integrate data on distances between fluorescent tags, assess the consistency of data collected by different experimental techniques, and link data available from techniques operating at very different length scales. Ultimately, all image information should be integrated into a dynamic computer simulation with molecular detail that will allow the investigator to observe, at the highest level of detail, the structure, assembly, and operation of the entire molecular machine and its interactions with other cellular components relevant to its functioning. This simulation should allow for changing the parameters of the system and predicting the resulting changes in function. Modeling of molecular dynamics on the time scale (nsec to msec) of many biological processes (e.g., protein refolding during complex formation) is an important, unsolved challenge.

An equally important mission of simulations will be the prediction of instrumentation performance. Details of the image-formation process will identify potential image artifacts and provide a basis for quantitative analysis of image imperfections including, for instance, the potential to correct for aberrations or AFM tip geometry. These simulations also will support design of advanced instrumentation with optimized imaging modalities.

Mechanical or radiative damage resulting from imaging may also be assessed through analysis of the process of damage followed by simulation. This analysis will allow for optimized data-collection and data-analysis strategies to compensate for the effects of damage.

Data Management

Any imaging techniques applicable to the study of molecular machines will result in the generation of huge amounts of data. Storage, manipulation, visualization, registration, and comparison of image information will represent a very

significant challenge. Separate images of identical complexes using either the same or different techniques will need to be compared and contrasted. Extracting information that will lead to functional insight from the vast amount of image data will require significant advances in visualization and image classification.

The creation of comprehensive image databases also represents a very significant technical challenge. Data intrinsic to images in one, two, or three dimensions, not to mention dynamic (fourth dimension) or correlation information (additional dimensions), are not readily organized in a universally applicable format. Instrumental and specimen parameters also will be essential parts of any useful database. A detailed analysis of image database needs should be addressed at the earliest possible time. Standardization requirements for these data protocols need to be defined early in the program. This will aid in the design of experimental protocols that will define the parameters necessary to maintain the long-term usefulness of the data collected. Public access to image-analysis software and databases of images will be an important product of the GTL program.

Because of the importance of computational tools and methods to the use of imaging in biological research, they are discussed in detail in Section 5 of this report.

Production of Lab-Scale X-Ray Sources

One high-impact development that would allow the scientific community to take full advantage of the use of X rays in imaging protein machines would be the creation of tabletop or lab-scale X-ray sources.

2. Intracellular and Cellular Structure, Function, and Processes

This section addresses needs for molecular imaging of protein machines and other key species within living cells. In response to GTL Goal 1, "Identify and Characterize the Molecular Machines of Life—the Multiprotein Complexes that Execute Cellular Functions and Govern Cell Form," the focus is on understanding the role of these machines in living systems.

Obtaining information on sub- or intracellular processes from living cells while they exert their normal function is highly desirable. These processes involve transduction pathways within a cell, cell reaction to external stimuli or signaling, and internal or external processes that lead to cell aging and apoptosis. To achieve such understanding, experimental challenges have to be met:

- Imaging techniques that provide an overview of cellular processes have to be reasonably fast, ideally on the order of a few images per second or faster. Imaging techniques with high resolution need to be noninvasive or minimally invasive so as not to perturb the cellular function.

- Probes or local reporters need to be developed that can easily be taken up by and have minimal effect on the cell. These probes also have to be extremely stable (neutral) to photon interactions and should adapt to cellular environments. An ideal probe would change, for instance, its property (and therefore report) changes in pH, oxygen concentration, or glucose concentration, while maintaining its response to external probing. Such probes should also be highly localizable within the cell and convertible to X-ray– and electron-dense markers.

Promising approaches to these problems are the newly developed vibrational microscopies and spectroscopies, X-ray microscopy, and electron microscope (EM) tomography.

Understanding the Genetic Basis for Microbial Function

Developing novel methodologies would be valuable in helping to establish the relationship between the genome and microbial function in living cells. First, several questions need answering: Can cell-imaging methods be devised to determine the direct correlation of gene mutations and gene expression to function in microbial cells? Can methods be developed to relate gene activation to the simultaneous detection of metabolism and secretion in microbial cells? Can high-throughput, live-cell imaging methodologies be formulated for approximate and selective genotyping? Studies of community dynamics under environmental perturbations would be greatly enhanced through real-time, genetic classification of individual microbial cells within a mixed culture, even if this classification is only approximate and must be further refined by more rigorous genotyping methods.

Issues and Limitations

The following will require technological advances:

- Imaging of single molecules (and sets of molecules) within living cells.

- Improved spatial resolution and optical probes for co-localization of data from different techniques below the resolution limits of standard optical images.

- Development of activity-neutral dyes and other optical probes.

- High-throughput approaches that enable the rapid analyses of large numbers of cells.

Capabilities developed for this element of the GTL program are expected to have wide applications to the study of cellular biology.

Recommendations

The following actions are recommended:

- Coordinate program to develop all methodologies, in particular co-localization of data from different techniques.
- Establish databases (e.g., protein spectra).
- Develop open source, common data formats.
- Stimulate partnerships between biologists and instrumentation scientists.
- Focus on specific needs of each methodology (e.g., some need more facilities, and others need stimulated development).

General Requirements

Recently, many promising techniques have been developed for studying intracellular and cellular structure and function, as well as the time evolution of cellular mechanisms via kinetics and dynamics. Ultimate requirements are capabilities for imaging functional groups and studying kinetics and dynamics in samples of a few microns in different environments, in vivo, and with a spatial resolution of tenths of a nanometer. Clearly, no one technique will offer the optimum in all required categories. Further, integration of data from techniques at different size scales is necessary. Real progress is likely to rely on multidisciplinary teams (e.g., physicists, chemists, and biologists) working together.

Current techniques can provide spatial resolution of a few tens of nanometers, with both chemical and elemental specificity and dynamical information on the picosecond time scale. Work on techniques that perform at this level is preliminary at many laboratories, and a database is needed to enable comparison of techniques and to provide benchmarks of technical progress in advancing particular imaging tools. Sample-preparation standards will allow comparable samples (and sometimes the same samples) to be measured by a variety of techniques for correlation of results.

The current evolution of techniques is toward higher spatial and spectral resolution, establishing data that, if properly organized and shared, can be used to identify similarities among the thousands of prokaryotic and eukaryotic organisms and their constituents.

Specific Needs

Most high-resolution (<50-nm) imaging techniques work well with thin (≤10-μm) specimens. Recently, however, cells were shown to function differently in 2D and 3D cultures. Consequently, an increasing number of scientists are growing cells in 3D matrices. Although these cultures more closely resemble the cell's natural environment, they present unique imaging challenges. Obtaining information about molecular interactions in these cultures without excessive and potentially damaging processing is extremely important. We need to develop tools that can collect data about the location of DNA, RNA, and proteins, as well as molecular interactions, at better than 10-nm resolution without having to process the cells (e.g., dehydrate, embed, and section the cultures).

We have the ability to obtain high-resolution images of whole, hydrated cells. We also can obtain exquisite information about the genetic and protein composition of cell populations through DNA and protein microarray analyses. Combining the power of these two technologies could enable the imaging of live cells and furnish information about the cell's genetic and protein composition.

Understanding the internal functioning of cells in their natural environments is important and particularly challenging. Analysis often requires the ability to visualize the cells within thick, dense matrices. For example, specific molecular (DNA, RNA, and protein), structural, and functional information must be obtained from specific bacteria within biofilms in their natural environment.

Powerful tools exist to study molecular machines in vitro as well as to obtain structural information about these machines from crystallography and STEM techniques. We need to be able to examine these machines within living cells to obtain functional information (in addition to that obtained using optical methods such as FRET) at better than 50-nm resolution.

Tools also exist to examine protein-protein interactions in living cells both in vitro and in situ using fluorescent imaging techniques. We need to be able to obtain structural and functional information about molecules (DNA, RNA, proteins, and protein-protein interactions) within living cells in living organisms at better than 10-nm resolution.

Methods are needed that will ultimately allow us to track the movement of macromolecules, vesicles, molecular machines, and other such structures within a cell with nanometer-scale precision. Current developments in this field are based on the detection and accurate position determination of bright labels such as semiconductor quantum dots, dye-loaded spheres, and light-scattering particles. Other attractive probe developments being made in the area of delayed fluorescence can detect single probes against the large background of cell autofluorescence by time-gating the photon collection.

Current Methodologies: Status and Development Needs

X-Ray Microscopy

X-ray microscopy today uses electromagnetic radiation with photon energies in the range from 250 eV to 30 keV, with corresponding wavelengths from 45 nm Å to 0.4 Å. Because of the short wavelength, suboptical spatial resolution can be obtained. Several different contrast mechanisms can be used to form an image, thereby making X-ray microscopes useful for probing different properties of matter. X-ray fluorescence mapping can be applied to mapping trace metals such as Fe, Cu, Zn, As, Se, Cd, Hg, Pb, or U in organisms including microbes or eukaryotic cells, while simultaneously collecting distribution from P, S, Cl, K, and Ca.

The critical element of any X-ray microscope is the zone plate used for imaging or focusing. The outermost zone width and aspect ratio define the achievable spatial resolution and diffraction efficiency. Therefore, a major investment in zone-plate technology is to improve the resolution. Achieving an outermost zone of 10 nm or less, while maintaining a thickness in the range of 200 nm for soft X rays and >>1 μm for hard X rays, would enable a major breakthrough in all applications. Another challenge is X-ray tomography. While transmission contrast imaging lends itself to 3D imaging in both the soft and hard X-ray region (Figs. 2.1 and 2.2), 3D imaging using X-ray fluorescence—even though potentially very powerful (Figs. 2.3 and 2.4)—is currently not feasible because of the long exposure times involved. A major improvement could be achieved by using novel detectors to maximize the solid angle of detection and by improving the brilliance of hard X-ray storage rings.

X-ray imaging, an emerging field, will be invaluable for studying hydrated organic matter at high resolution, particularly organic and inorganic composites such as bacterial colonies (e.g., in soil, microbial communities, and biofilms). X-ray tomography can be carried out on whole cells or colonies of bacteria up to 10 μm thick, without the time-consuming sectioning required for TEM, and generate a 3D reconstruction of the cell or colony within 1 h at ~50-nm resolution.

In addition, the entire arsenal of heavy-metal labeling methods for TEM can be used for X-ray microscopy. As a result, correlated live-cell light

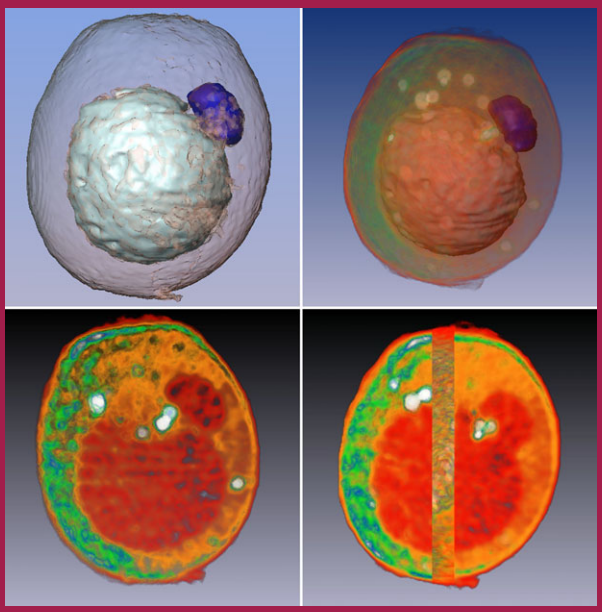


Fig. 2.1. Three-dimensional soft X-ray reconstruction of yeast. Example of 3D volume visualization of a tomographic data set showing various pixel- and surface-classification techniques. [Source: Carolyn A. Larabell and Mark A. LeGros, Lawrence Berkeley National Laboratory]

and high-resolution X-ray imaging analyses can be used to examine both the dynamics of specific proteins and the ultrastructural distribution of those proteins in the same cell.

Future implementations of mixed 3D imaging involving X-ray, zone-plate direct imaging, and lensless diffraction imaging promise to extend the spatial resolution of 3D X-ray methods to the nanometer level in thick, hydrated samples. X-ray imaging will, therefore, play a unique role in quantitative 3D measurements of protein and elemental distribution at high spatial resolution.

Optical Microscopy

Optical methods vary in complexity, from inexpensive and routine tools to major interdisciplinary investments at the national laboratories. The accessible distance scale ranges from the "optical limit" of about half the wavelength to macroscale or, in other words, from the bacterial and cellular organelle scale to sample sizes limited only by optical transparency. Image production providing rich information content is a result of a wide variety of physical and chemical properties. In favorable circumstances, the detection limit may be a single fluorescent molecule. We can classify optical imaging methods according to difficulty and application to define a role for GTL in meeting the needs of the biological community. This classification also is intended to show where multidisciplinary teams of physicists, chemists, and biologists should work together to produce the capabilities required for the next generation of methods and instruments. Each of these methods is discussed in some detail in Appendix C.

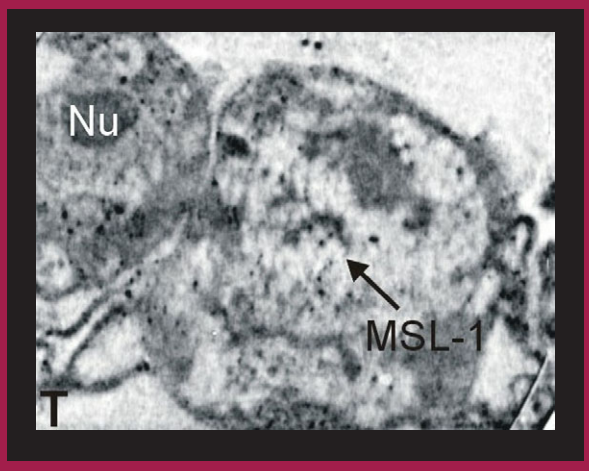


Fig. 2.2. Soft X-ray tomography. Section through a tomographic reconstruction of a *Drosophila melanogaster* cell. Cells were labeled against the MSL-1 protein using 1-nm colloidal gold conjugated antibodies and silver enhancement. Cell nucleolus (Nu) and MSL-1 distribution positions are marked. [Source: G. Schneider et al., "Nanotomography of Labelled Cryogenic Cells," *SPIE Proceedings* **4503**, 156 (2001)]

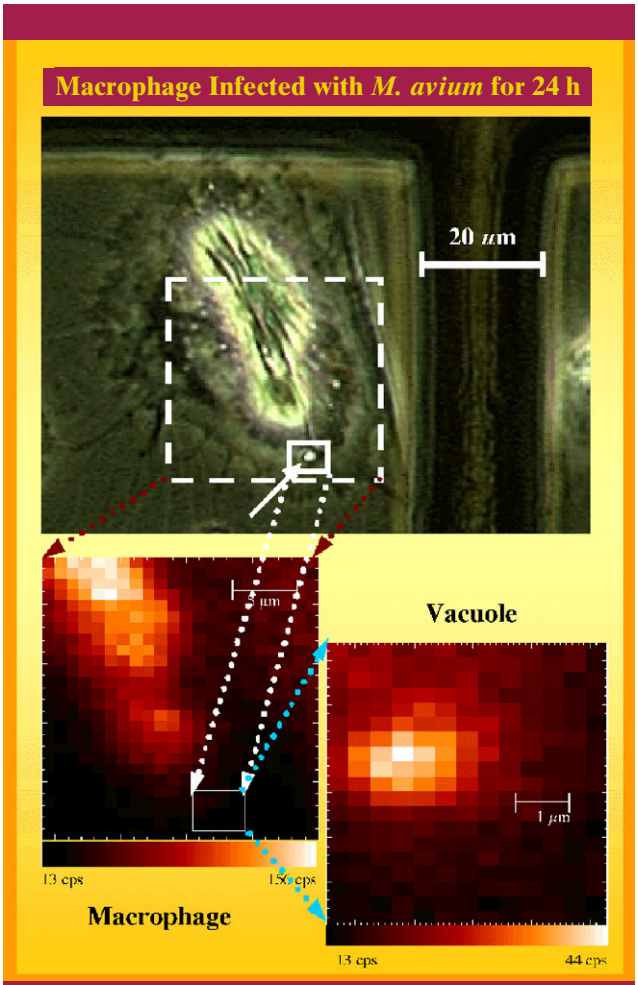


Fig. 2.3. X-ray fluorescence image of *Mycobacterium avium*, 24 h after infection of a fixed and air-dried macrophage. A map of the Cl distribution of both macrophage (30- by 30-μm field) and mycobacterium (3- by 3-μm field) is shown. [Source: Dirk Wagner, Kuzell Institute for Arthritis and Infectious Diseases; Jörg Maser, Barry Lai, and Zhonghou Cai, Argonne National Laboratory; and Luiz Bermudez, Kuzell Institute for Arthritis and Infectious Diseases]

Far-Field Imaging Methods Based on Intramolecular Vibrational Modes

Far-field vibrational-imaging methods will complement fluorescence- or absorption-based optical approaches for the study of individual cells and colonies. The use of synchrotron radiation and lasers provides diffraction-limited spatial resolution that could be improved further by image-enhancement algorithms. Vibrational signatures specific to molecular structure can provide direct chemical information without the use of dyes or other labels. In addition, vibrational imaging can be exploited to observe dynamic changes in living cells. Far-field infrared vibrational imaging can be developed to examine community signaling or colony and substrate chemistries at the scale of 10 μm and greater.

Collectively, vibrational-imaging techniques should be developed to analyze the microbial response, at the level of individual cells or communities, to perturbations such as changes in oxygen or other substrate levels (e.g., NH_3 and metals), pH or temperature changes, and other factors that will affect critical elements of the interaction of microbes with their environment. Because such hyperspectral-imaging methods can potentially yield large data sets including spectral, spatial, and temporal dimensions, imaging-analysis methods that can properly exploit and manage such data sets should also be developed.

In addition to their application to intrinsic molecular signatures, vibrational-imaging methods could be developed to take advantage of the introduction of vibrational "dyes." Thus, "tagging" the pathways or structures in microbes by the use of vibrational labels such as stable isotopes or ligands (e.g., CN and CO) could allow selective detection of specific molecular components within and among microbes.

Near-Field and Scanning-Probe Imaging Methods

Near-field microscopies typically are based on the spatial confinement of photons by probe microscopy structures. They offer attractive pathways to achieving resolution beyond the diffraction limit and potentially could combine topographical imaging with chemical sensitivity. Because of their dependence on probe microscopy techniques, near-field probes are usually limited to cell exteriors, but this could be overcome with newly developed photonic force microscopes that could "grab" artificially introduced labels and move them within the cell to probe the interior.

Chemically functionalized near-field tips also could be injected into the cell through the membrane and positioned to measure the uptake of chemicals through the cell membrane or their diffusion within the cell. The development of near-field imaging capabilities that could use narrow-band laser light to achieve λ10 with apertures and eventually λ100 without apertures with scanning probes in living cells and membranes is highly desirable.

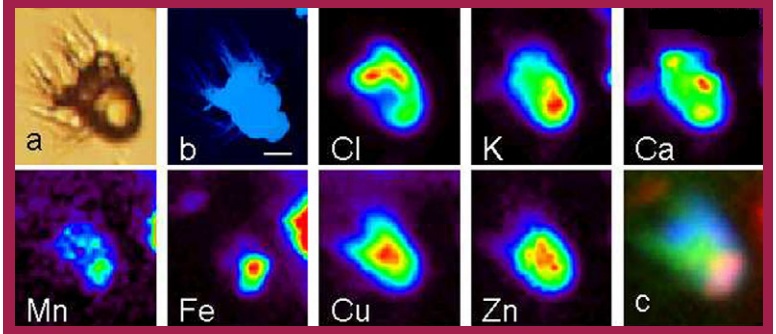


Fig. 2.4. (a) Soft X-ray micrograph and (b) X-ray fluorescence map of marine ciliate showing the cell's C content (scanning X-ray microscope, National Synchrotron Light Source at Brookhaven National Laboratory). [Scale bar (on b) = 6 Fm] Cl, K, Ca, Mn, Fe, Cu, Zn: X-ray fluorescence maps showing distribution [hard X-ray microprobe, Advanced Photon Source, sector 2, at Argonne National Laboratory (ANL)]. c: Superimposition of Cl, Fe, and Cu maps. Cl was chosen as a proxy for cellular biomass, while Fe and Cu represent two trace metals with notably different spatial distributions within this cell. [Source: Benjamin S. Twining, Stephen B. Baines, and Nicholas S. Fisher, Stony Brook University; Chris Jacobsen and Jörg Maser, Argonne National Laboratory]

"Classical" One-Photon Methods

All classical microscopy methods use continuous or low-power light sources. Contrast may be improved by physical methods such as phase or differential-image contrast. Specific and very useful information is provided by fluorescence methods that use intrinsic or added dyes. In general, fluorescence techniques and many fluorescent probe molecules are available, but improved labels are needed.

Obtaining chemical data (location of abundant species or important minor chemical compounds) is difficult with the above techniques. A direct means for obtaining molecular information is by vibrational resonances, either infrared or Raman. Even though improved detection technology will be useful, technologies for Raman and fluorescence measurements are approaching their theoretical limits, requiring the development of improved probes.

The now-accessible terahertz spectral range yields many vibrational excitations that can be explained by molecular dynamics calculations. High-power sources available at the national laboratories can be used for imaging applications, including detection of bioterrorism agents such as anthrax. Additional research is required to achieve the potential of this approach.

A special imaging category is scanning optical tomography, which requires a pulsed light source to provide depth profiling data. Although the laser must produce ultrashort light pulses, high power is not required, and obtaining 3D information useful on a tissue scale is possible, but not at the single-cell level.

Multiphoton Microscopy

The most promising avenue for obtaining new kinds of cellular information—and the area where funding at the national laboratory level can provide the greatest impact—is in multiphoton methods. A wide variety of characteristics can be studied, with the general requirement of an ultrafast laser technology. Multiphoton methods, which show real promise for providing chemically based imaging, offer the unique advantage of possibly generating several imaging signals simultaneously. All signals are generated at the laser focal point and can be separated to create a rich, perfectly registered, multidimensional-image data set containing a variety of chemical, physical, and structural data. Wider access to multiphoton methods will require improved (smaller, cheaper, and more reliable) laser technology and optical probes. Advanced imaging tools will help continue progress in solving a wide variety of biological problems.

Hole-Burning Microscopy

A hole-burning imaging (HBI) approach for probing local environments in microbial cells can be viewed as an optical analog of MRI. This high-sensitivity technique has the potential for detecting local effects on molecular probes. In principle, HBI has the spatial resolution of multiphoton microscopy methods (i.e., on the order of 100 nm). Whereas MRI is based on proton T_1-relaxation times, HBI is based, in part, on T_2*-pure optical dephasing times of organelle-specific dye molecules. The sensitivity of T_2* to subtle differences in molecular environment gives the technique its high specificity.

Dynamical Studies

Direct measurements of kinetics and dynamics in living cells could be obtained on a picosecond time scale by adapting pump-probe methods to obtain high spatial resolution. The nanosecond time scale can be probed by measurements of lifetimes and rotational correlation times that are sensitive to processes such as protein association to form complexes. Approaches of this nature have provided valuable information on a variety of biological systems, for example, retina function, photosynthesis, and energy conversion.

Electron Tomography

Electron tomography has the potential to reach the highest spatial resolution in imaging 3D structures of intact cells and mapping subcellular component locations. Theoretical analyses suggest achievable resolution of 2 to 5 nm with such specimens. At this resolution we should be able to identify such subcellular features as cytoskeletal filaments and nucleic acid condensates and such large macromolecular complexes as ribosomes, proteasomes, and polymerases. This could be achieved by matching the observed

densities with structures of corresponding molecules that have been solved by X-ray crystallography or other techniques.

Although being able to directly identify the larger protein complexes will provide a wealth of new information, selective labeling of components with electron-dense agents will enable localization of more components. The labeling strategies could include conventional antibody-targeted metal atom clusters. The use of quantum dots, for example, would allow correlative light microscopy, but the particle size for electron microscopy would be smaller than for current light microscopy.

One problem to be addressed is how to get the particles into the bacteria. Evidence suggests that complexing them with a naturally imported peptide such as transferrin may work. Other labeling approaches under development include modifying a particular gene to include a photoconvertable eosin or tetracysteine peptide that can be complexed with a heavy metal atom cluster.

Instrumentation development probably will be driven by the materials science community. Specimen preparation for frozen-hydrated, whole-mount samples is fairly well developed, but improving the methodology for frozen-sample sectioning would open up the technique to applications with multicellular samples. Finally, managing the image data sets and searching the reconstructed volume densities for patterns corresponding to known structures will require significant computational efforts.

3. Monoclonal and Heterogeneous Multicellular Systems, Cell-Cell Signaling, and Model Systems

SECTION

Goal 3 of the GTL program is to "characterize the functional repertoire of complex communities in their natural environments, at the molecular level." Even though the genomic sequence for several different organisms is known and we understand the general outlines of how some communities are composed, we lack an understanding of community function distribution among members and the molecular basis of different functions. This information is required to connect molecular expression data to the overall function of a community in a natural setting. To bridge this enormous gap of complexity, there is a need for simple, well-characterized model systems (i.e., laboratory, bench-scale, pure-culture, and mixed-culture cell systems) that capture essential functional aspects of natural communities (Fig. 3.1).

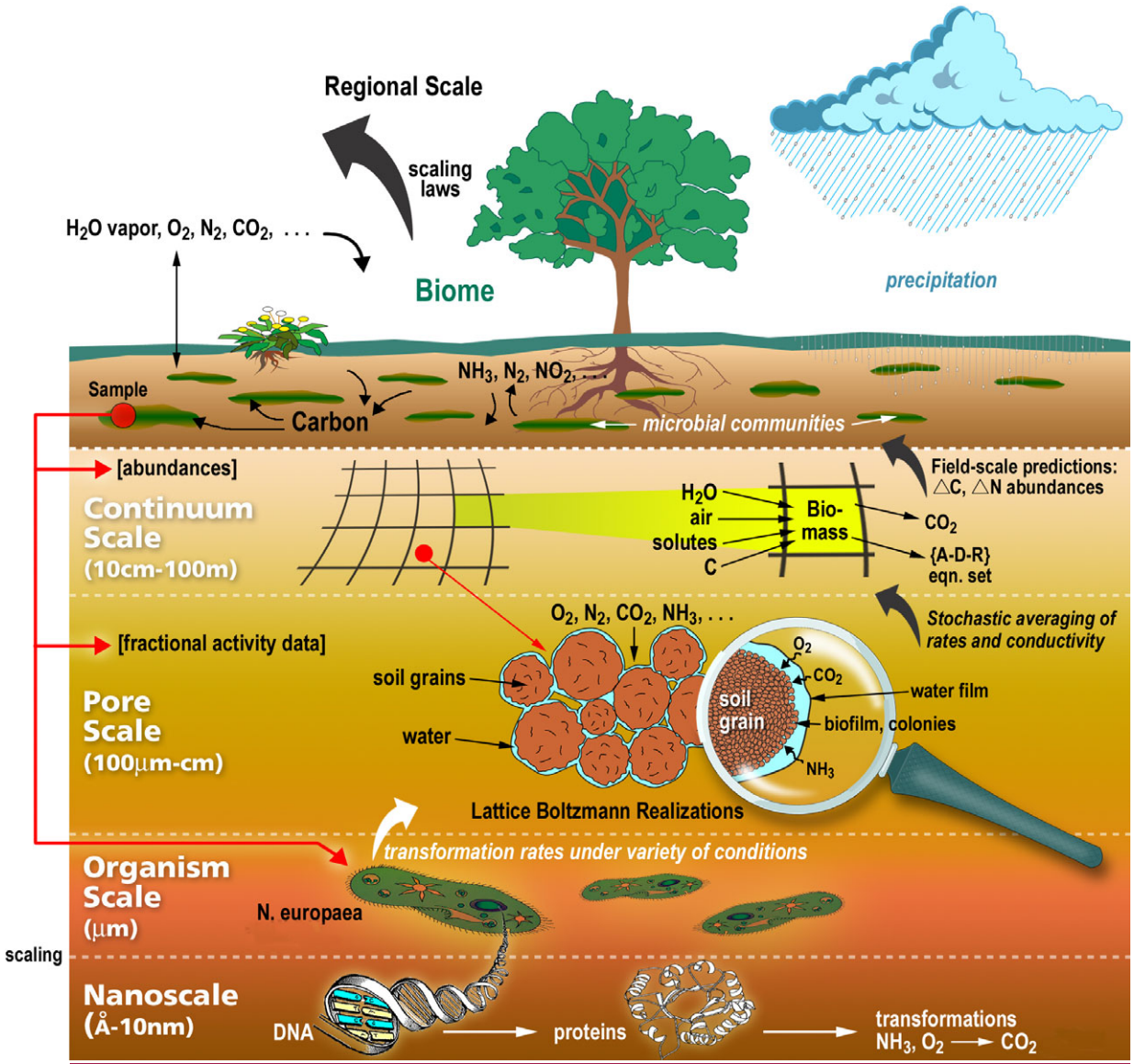


Fig. 3.1. Relevant level of complexity inherent in microbial soil ecosystem.
[Source: Pat J. Unkefer, Los Alamos National Laboratory]

Why Imaging is Needed

Imaging technologies are ideal for capturing the complexity of multicellular interactions because of their ability to acquire spectral, spatial, and temporal data over a wide range of scales. Because of their parallel and quantitative nature, imaging approaches are inherently high throughput (i.e., they can accommodate thousands to tens of thousands of simultaneous experiments). To realize this potential, model systems must be adequate for developing and validating imaging technologies. In turn, well-defined, multicellular model systems can reveal important principles that govern cell communities. Both monoclonal (pure-culture) and heterogeneous, multicellular (mixed-culture) systems need to be defined. Specifically, we must identify

- Structural and functional properties of multicellular systems,
- Imaging techniques needed to determine these properties, and
- Major technological challenges.

Understanding complex communities of cells, including how cells interact with their own and different species, is fundamental to the goals of the GTL program. Interactions include the exchange of information through signaling mechanisms involving chemical, biochemical, and ionic mediators. These processes, in turn, are directly impacted by the extracellular environment. Thus, understanding the interaction between various cells and the environment is essential for understanding complex communities.

Imaging technologies can provide this understanding; the challenge, however, is to image information flow among cells. First, we must understand the architecture of the molecular machines that mediate such flow. We must determine which cells are talking and which are listening. To do this, we must be able to measure signaling dynamics and image metabolites and regulatory molecules. In addition, we must be able to track metabolites and regulatory molecules spatially and temporally. Finally, we must connect this information to the consequent cellular responses (Fig. 3.2).

Imaging Issues

Several technological issues must be addressed, the first being those of spatial and temporal resolution. Microbes and organelles are close to the theoretical resolution of optical microscopy. Signaling molecules and metabolites are normally found at exceedingly low concentrations. Natural microbial communities are frequently found in optically opaque conditions or in the presence of interfering compounds. Experimental model systems can be constructed to prevent some of these problems, but they give rise to other issues. For example, high diffusivity in culture systems can affect sensitivity and spatial resolution. Then there are the fundamental problems related to measuring signal-dependent cellular responses and identifying unlabeled signaling molecules.

Improved Model Systems for Microbial Imaging

In the context of the GTL program, several scientific questions have been identified (see other sections) that are relevant to the goal of understanding the molecular basis of complex cell communities. Approaching these questions in

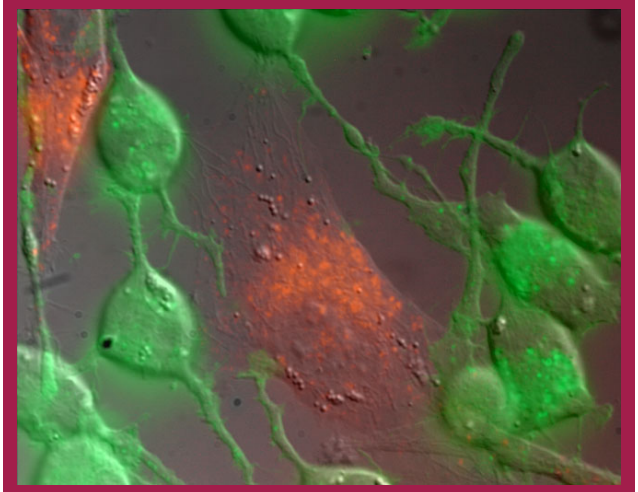


Fig. 3.2. Visualization of cell-cell communication in living cells. Mouse cells were engineered to express either a membrane-anchored growth factor tagged with a red fluorescent label or the complementary receptor tagged with a green fluorescent label. The cells were visualized by differential interference contrast optics. [Source: H. Steven Wiley, Pacific Northwest National Laboratory]

the context of natural communities is very difficult, however, and thus should be initially studied in well-defined model systems. In addition, a variety of powerful experimental imaging methodologies currently can be used only on model systems, and some cellular processes occur only in a multicellular context. We should use model systems that are simple enough to facilitate the development of imaging technologies but are still relevant to complex environments. The specific characteristics and components of an ideal or nearly ideal experimental system include

- Sequenced genomes,
- Genetically tractable genomes,
- Cultures grown and maintained under defined conditions,
- Culture state followed by measurable parameters, and
- Culture geometries suitable for multiple imaging technologies.

Specific scientific questions will, in large part, drive the choice of experimental system. For example, monoculture systems are most appropriate for investigating quorum factors, whereas mixed-cell systems will be required for investigating metabolic coupling, such as that observed in nitrifier (nitrate-converting), denitrifier (nitrate-removing), and syntrophic[1] systems. The challenges of monoculture vs mixed-culture systems are distinct, and both types of model systems should be developed.

Challenges in producing multidimensional model systems include culture stability and composition, analytical techniques appropriate for complex systems, and individual cell manipulation. We need to define how multispecies cultures evolve. How do we track all the parameters that change? Is there a way to create a virtual "chemostat" for multicellular systems? Both symbiotic and syntrophic systems are particularly attractive because of their mutual dependence, which will allow identification of temporal and spatial regimes where imaging can be productively used. Model systems should be standardized to facilitate information interchange.

[1]Syntrophic associations include two or more organisms for which intimate interaction is required for the complementary exploitation of resources.

Interrogation of Model Systems

Imaging technologies can provide very valuable information on the dynamics of cellular populations and the molecular processes occurring with the cells. To link molecular and cellular properties to overall community function, a number of parameters must be followed simultaneously. Imaging-based approaches should follow as many parameters as possible to create a high-throughput research capability using sample-preparation techniques that will not perturb the chemistry of the cells or their environment.

Identify Different Cell Types and "State"

Specific ligands and probes for cell labeling, both gene and protein level, are needed. We need to better define unique spectral signatures of specific cells and their metabolic state and associate cells with specific gene-expression patterns; thus specific marker genes should be identified and this information used to create "reporter" cells by inserting fusion proteins behind specific genetic elements. Better probes to directly monitor the metabolic state of cells are also needed. Advantage should be taken of endogenous cofactors that display absorbance spectra dependent on metabolic state. Visualization approaches should be developed to identify the activation state of specific signaling networks triggered in response to environmental perturbation.

Track Individual Cells in Time and Space

There is a strong need for both high-resolution and wide-field technologies. Being able to quickly switch between these two modes would be extremely valuable because it would allow tracking of both cell communities and individual cells. Following the trajectory of individual cells over a significant length of time, sometimes for many days, is often necessary. Useful model systems will include both 2D and 3D geometries; therefore, cell tracking should not be restricted to any particular one. Following individual cells for significant lengths of time also requires the development of stable cell-identification probes or establishment of specific spectral signatures.

Track, Quantify, and Identify Signaling Molecules in Time and Space

Another strong need is to define spectral signatures of important signaling molecules or to develop appropriate "tags." A valuable approach would be to make "detectors" or optical sentinels that can be placed in the extracellular environment. Other needs are specifically labeled signaling molecules and tagging approaches that allow correct transport with high signal-to-noise ratios. Obviously, these approaches should not interfere with function. Additional attention should be devoted to developing infrared, X-ray, nuclear magnetic resonance, mass spectroscopy, and Raman measurements of metabolites both within cells and in the extracellular environment. Nondestructive nanoscale probes could provide information without interfering with imaging technologies, and many of these probes could circumvent limitations of optical techniques.

Understand Cell-Environment Interactions

To understand multicellular systems, we also must investigate interactions between cells and their physical environment. Different physical and chemical interfaces will have distinct effects on individual cells within the community that can, in turn, influence their interactions. Effort should be focused on defining important environmental aspects that impact specific multicellular systems.

Many of the same techniques for quantifying metabolic products can be used to study the interaction of cells with their extracellular environment, but additional approaches will be required. Examples include (1) imaging technologies that can be "tuned" to specific classes of compounds in the extracellular environment, such as metal and organic compounds and reactive species and (2) specific probes for such identified compounds as organic molecules found in extracellular polymers, secreted proteins, nucleic acids, polysaccharides, metals and their states of oxidation, chlorinated aromatics, and aliphatics. In particular, we need to image species that are present, concentration gradients associated with chemical potential, and different pH states. Methods for visualizing macroscopic transport processes also should be developed.

Understanding what cells produce to modify their environment is important. Efforts should be made to determine all the proteins secreted by cells or actively shed from the cell surface, including both soluble and insoluble proteins. Because environmental modification will vary with time, labeling techniques are needed for examining rates of synthesis and deposition of proteins and other organic molecules. These techniques can include isotopic and fluorescence labeling and other suitable imaging approaches.

Required Instrumentation and Data Analysis

Imaging instrumentation for multicellular systems should have the following ideal characteristics:

- **Noninvasive.** Nondestructive, with high temporal and spatial resolution.
- **Highly sensitive.**
- **Multimodal.** Able to span multiple spatial scales and see population behavior (wide field of view), then zoom down to an appropriate resolution (e.g., molecular). For example, use one mode for identifying a particular cell and then use a high-resolution mode to specifically interrogate the cells or simultaneously gather multiple data streams for temporal and spatial correlations that can establish functional relationships.

Because of the complexity of cellular processes, establishing cause-effect relationships is often impossible without precise correlative data. Instruments should be accessible to many different groups of investigators. Accessibility can be guaranteed either by reserved time at user facilities for imaging or through the availability of onsite culture facilities. A strong need also exists for development of compact (bench-scale), affordable instrumentation for single- or multilaboratory use. Other needs are improved software for automated image acquisition, analysis, and customized statistical analysis and new algorithms for multivariate analysis and multimodal imaging.

4. Imaging Microbial Communities

Microorganisms and the processes they catalyze are fundamentally important to DOE missions in energy and environment. Microorganisms rarely live alone in nature but rather in populations and communities of varying complexity. These communities drive the flow and cycling of energy and elements (biogeochemical cycles) in the biosphere and contain most of the biochemical diversity on earth.

A major goal of the GTL program is to develop a better understanding of the structure and function of microbial communities from their genome sequence–based information. Imaging can play a major role in testing hypotheses and predictions generated from this information at the community level. The understanding of community structure and function will enable the manipulation of microbial communities for the benefit of mankind. This includes meeting such DOE needs as environmental cleanup, carbon sequestration, global climate change, and clean energy production. The advancement of imaging for systems of this scale and complexity will also advance DOE's health-related research in low-dose radiation and radiation therapy.

Issues

DOE's Microbial Genome Program continues to produce whole-genome sequences that are being used to study the detailed biology of dozens of individual cells and populations of microorganisms of interest to DOE. The GTL program is providing extensive genome sequence–based data on microbial communities, but a major challenge is to link all this information to communities in nature. Structural and functional analysis will be needed because microbial environments can be extremely diverse and can influence the way communities function. Critical questions to be addressed include the following:

- How are the community and its physical environment arranged? Communities may be layered or composed of heterogeneous populations within biofilms. Physical and chemical properties in communities can vary dramatically over the scale of a few microns.
- What are the members, and where do they reside? What are their phylogenetic and functional properties?
- How are microbes interacting among themselves and with their environment? How does information (e.g., genes and signaling molecules) flow within the community? What are the ecological interactions such as predation and symbiosis among the various members?
- What is the flow of energy and elements through the communities, and how is this flow regulated?
- What are the intrinsic biological (genetic) and environmental factors that control the structure, stability, and functioning of communities?

Studying Microbial Communities

Most approaches currently used to study microbial communities are poorly suited for answering these questions. They provide, at best, a snapshot of a fraction of the community. Methods include culturing, bulk chemical, physical, and even molecular-based techniques that provide limited information on the spatial and temporal events critical to understanding community function. For example, information on gradients and spatial heterogeneity is lost in bulk analyses. The availability of whole-community genome sequences will provide an enormous amount of information about potential gene function and, therefore, community function. However, gene sequences will allow only function prediction, and such predictions must ultimately be tested

through in situ studies of communities. Microbial communities will require coupled biological, chemical, and physical analyses at multiple scales that may best be achieved through imaging.

Imaging, as it pertains to community analysis, provides spatial and temporal information on the chemical, physical, and biological characteristics of the community and its environment. This information can encompass length scales from the molecular to that visible with the human eye. Revealing this intricate structure provides insights into community function unavailable from bulk analyses. Further, imaging will provide the data required for modeling and testing predictions based on genome sequence–based information.

Challenges

Critical to addressing key scientific questions will be the novel application of existing and emerging technologies. Further, a number of scientific needs will require fundamentally new developments in imaging technology. The following are some of these needs:

- Spatial and temporal concentration measurements of signaling molecules and metabolites at the intra- and extracellular level and within populations and communities.
- Assessment of the role of extracellular polymers in community structure, function, and stability.
- Determination of macroscopic transport of water, solutes, and macromolecules and their relationship to microbial function.
- Detection and frequency of genetic exchange, recombination, and evolution within communities.
- Measurement of the elemental distribution, oxidation state of elements, and biomolecules within and among communities.
- Development and validation of models of community growth, function, and response to environment.
- Development of advanced chemical and biological probes, including engineered microorganisms, tagged biomolecules, and chemical sentinels that will help characterize microbial communities.
- Characterization of the physical and chemical properties of interfaces.
- Development of advanced tools for imaging characterization for use in the laboratory as well as in the field.

Specific examples of technical developments that would enable approaches to these challenges follow.

Specific Challenges

Biofilms and microbial communities can facilitate the formation of new solid phases (biomineralization) and interact with existing solids and with other microbes and soluble constituents in the environment. Determining the spatial distributions and gradients of elements and their redox state, as well as the distributions and gradients of molecular products (proteins and metabolites), can provide information required to better understand (1) how the microbial community and physical environment are arranged, (2) how microbes interact among themselves and with their community, and (3) the flow of energy and elements through the community and how that flow is regulated. Techniques are needed to image the complex 3D microbial communities and their environments and characterize the physical, chemical, and biological interactions occurring in them. Several currently available techniques need improvement. Some are minimally invasive or destructive, while others are noninvasive. Examples of such minimally perturbing techniques are given below.

Advanced Optical Methods

Optical spectroscopic methods can be used as tools for noninvasive characterization and monitoring of dynamic behavior. Measurements of absorption and in vivo fluorescence can be used to monitor the presence and relative concentration of optically active biochemical species. Light-scattering spectroscopy can probe the size distribution of structures in a community. Vibrational (infrared and Raman) spectroscopy is a technique for studying the composition of biological materials without perturbing or labeling the sample. Biological components (e.g., lipids, proteins, nucleic acids, and carbohydrates) and

biofilm and microbial surfaces (e.g., minerals and polymers) have unique vibrational spectra based on their chemical structures. The use of these methods will provide new information on the following:

- Large-scale (1- to 10-μm) biochemical organization.
- Composition and distribution of extracellular polymer matrices.
- Concentration and distribution of nutrients, metabolic proteins, signaling molecules, and other macromolecules.
- Interactions of biofilms and microbial communities with supporting surfaces.

Because vibrational spectromicroscopy is noninvasive, it can be performed on dynamic living systems in combination with other techniques. If synchrotron radiation is used as the photon source, a dynamic system can be studied directly on surfaces of geologic materials (Fig. 4.1).

Significant progress has already been made using confocal and two-photon fluorescence microscopy. The specificity of these techniques is provided by the exogenous chromophore. The resolution is on the order of a micron and is slightly higher for two-photon than for confocal microscopy. Delivering chromophores to remote regions is a particular challenge. Additionally, the identification of probes that maintain activity in diverse environments is required (Fig. 4.2).

All these techniques can be used in an imaging arrangement to monitor changes in community behavior in real time. Improvements are needed in such areas as spatial resolution, the ability to provide quantitative information, and data-acquisition speed. Additionally, advanced light microscopy techniques such as surface-plasmon resonance, surface-enhanced Raman spectroscopy, second-harmonic-generation imaging, optical-coherence tomography, and coherent antistokes Raman scattering microscopy can be developed for high-resolution 2D and 3D mapping, often with specificity to particular components associated with the imaging technique.

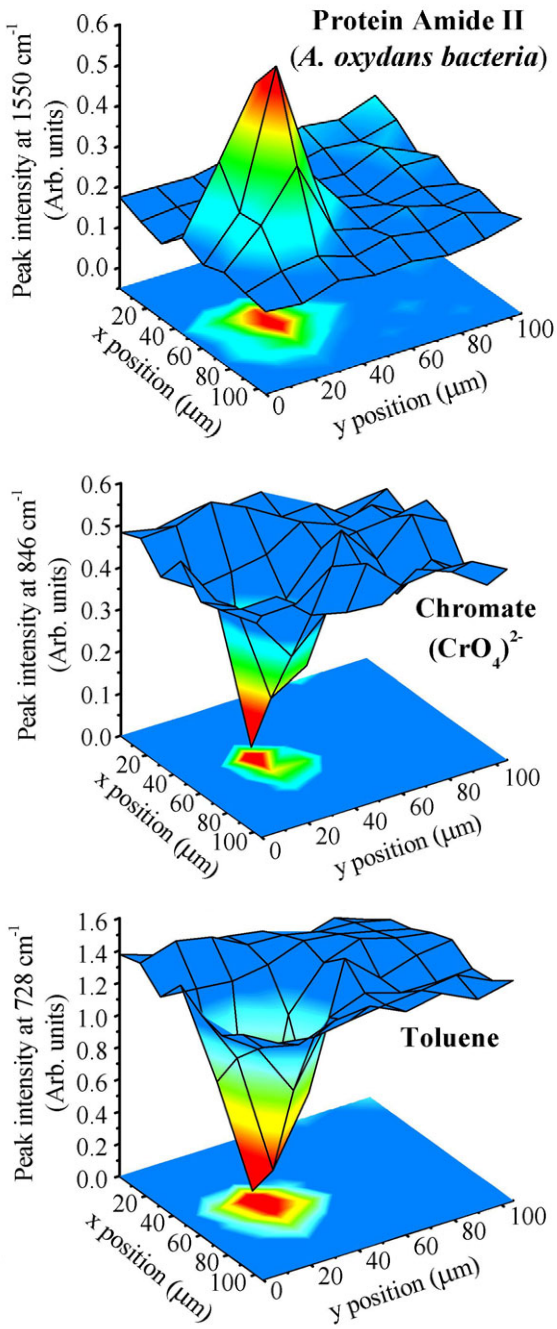


Fig. 4.1. Synchrotron infrared images showing a small colony of natural bacteria. A common organic contaminant, toluene (bottom), is used to accelerate the reduction of a carcinogenic form of chromium [chromate $(CrO_4)^{-2}$, middle] to its environmentally safe form. Bacteria are located via their spectral signature (top). [Source: H.-Y. Holman et al., "Real-Time Characterization of Biogeochemical Reduction of Cr (VI) on Basalt Surface by SR-FTIR Imaging," *Geomicrobiology Journal* **16**(4), 307–24 (©1999). Reproduced by permission of Taylor & Francis, Inc. (www.routledge-ny.com)]

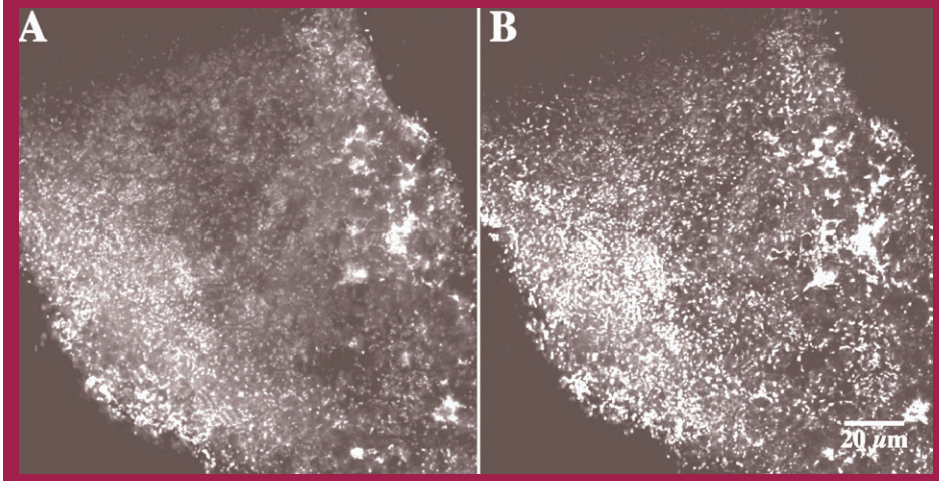


Fig. 4.2. : One-photon (left) and two-photon images of the same location and depth, showing a laboratory-grown biofilm stained with a nucleic acid–specific fluorochrome. Note the higher resolution achieved by two-photon excitation. [Source: Thomas R. Neu, UFZ (Centre for Environmental Research), Magdeburg, Germany]

Nuclear Magnetic Resonance (NMR)

Energetics and cell communication in a microbial community are governed by the flow of nutrient, waste, and signaling molecules to and from the cells. For numerical simulations, transport rates are often treated as constants even though they are known to vary with biofilm density. Imaging techniques are needed to image substrate and metabolite concentrations, biofilm density and permeability, and mass transport in biofilms and the surrounding (or supernatant) medium. Such techniques must be able to measure transport everywhere in the sample at the highest attainable resolution and under both quiescent and flowing conditions. They should be noninvasive and employ no added tracers. Furthermore, the measurements should be applicable to samples with special environmental requirements such as anaerobic systems. Finally, the techniques should allow imaging and comparison of the various parameters mentioned above on both a small sample (ideally a single microbe) and a larger microbial community and in their natural environments such as soils.

Nuclear magnetic resonance imaging (MRI) methods appear to meet many of these requirements. Existing MRI contrast techniques provide a spatial mapping of the sample: (1) structure via its hydrogen density; (2) molecular physical environment via NMR *spin-lattice* and *spin-spin* relaxation times; (3) chemical content via spectrally selective and chemical-shift imaging methods; and (4) fluid transport information via flow and diffusion imaging, all without opacity losses.

Fluid flow and diffusion imaging have been successfully applied to map convective and diffusive transport in biofilm and surrounding media. Improved NMR flow-mapping methods might integrate velocity and diffusion mapping, intracellular-extracellular water-exchange measurements, biofilm-density mapping, and metabolite-concentration mapping for comprehensive analysis of transport in planar biofilms and granular communities. MRI also can be combined in one instrument with optical confocal imaging to provide simultaneous data from both modalities.

Challenges for applications of this technique for in situ microbial studies include the development of novel MRI equipment optimized for studies of microbes in their natural environment; enhancement of NMR sensitivity; improvement of the spatial, spectral, and temporal resolution of the images; and development of multivariate data-processing methods.

Novel 3D Imaging Techniques

There are a variety of 3D imaging techniques, each with its own unique set of strengths and weaknesses (i.e., information type, contrast mechanisms, sample geometry requirements, spatial resolution, sensitivity, and probes). These techniques meet many but not all of the information requirements for a comprehensive characterization of microbial communities. Additional imaging techniques that provide complementary information—for example, acoustic imaging—are required to fill in the gaps. Further, resolution

enhancement (where feasible) is needed for some existing imaging methods to properly apply them to the fine scale of bacterial cells.

One approach is to identify and integrate complementary imaging modalities to attain a synergistic multimodal imaging capability. To support such combined imaging technologies, image-fusion methods for combining diverse image data with differing spatial resolutions are needed. An example would be optical fluorescent methods combined with MRI. A second approach is to study a specific microbial process with all available imaging techniques, combined with comprehensive data analysis using advanced computational techniques. This integrated method should improve understanding of the links between gene sequencing of individual bacterial cells and biofilm functioning in the environment.

Microsampling Techniques

In some cases, relying on direct determination of spatial and temporal distributions of chemical and physical properties will be necessary. Reasons include large background, lack of an appropriate contrast mechanism, lack of a probe or contrast agent, and unacceptably large perturbations to the system caused by the agent and low levels of species to be mapped. In these instances, instruments are required that can provide detailed, high-sensitivity maps of chemical species and physical properties.

Effective microsampling techniques must interface with analytical techniques in studying microbial communities. Microsampling can be used to precisely extract microorganisms or chemical samples for subsequent separation and analysis. Various analytical techniques can provide biosignatures. Advanced analytical tools that provide rapid chemical and biological information are needed. This combination of microsampling with analytical techniques will allow the imaged microbial behavior and functions to be related to specific metabolites, protein profiles, and genetic materials of microbial communities.

Advanced Chemical and Physical Sensors

Sensors, by definition, are instruments that detect and quantitate one or more agents or properties in the presence of the surroundings. To directly address the large issue of background, current-generation sensors either exploit a unique signal or characteristic or, more commonly, restrict the spatial region or separate the desired agent from its milieu.

Future sensors, to be of use in the characterization of microbial communities, must integrate the spatial segregation-separation function with the detection-quantitation function; therefore, a new generation of instrumentation must be invented to handle sophisticated sequential and parallel chemical manipulations simultaneously with ultrasmall sample quantities. Critical functions in microbial communities most likely will be implemented with signaling molecules that may be present at small copy numbers within any given cell. Thus, new-generation sensors must be capable of previously unavailable limits of detection. Furthermore, even though current-generation sensors address at most a few analytes, massively parallel measurement strategies must be invented to simultaneously measure a large number of different species with high spatial and temporal fidelity.

Finally, it will be important to understand how sensors affect the properties of the microbial community, especially physical sensors whose contact with the community may be part of the measurement operation. When measured against these requirements, the chemical and physical sensors of today are slow, insensitive, poorly integrated, and capable of very limited sample manipulation. A key goal will be to advance the state of the art among these figures of merit to make significant contributions to understanding microbial communities.